DIAGNOSTIC ACCURACY AND SHORT-TERM SURGICAL OUTCOMES IN CASES OF SUSPECTED ACUTE APPENDICITIS

Shi Wu Wen, MB, PhD; C. David Naylor, MD, DPhil

Abstract • Résumé

Objective: To test the hypothesis that, with modern diagnostic methods and antibiotics, more conservative use of surgery in cases of suspected appendicitis would not result in increased rates of short-term complications in confirmed cases.

Design: Population-based observational study using administrative data.

Setting: All Ontario hospitals in which primary appendectomy was performed from Apr. 1, 1981, to Mar. 31, 1992.

Patients: All 126 815 patients admitted to hospital for a primary appendectomy during the study period.

Outcome measures: Diagnostic accuracy rate (acute appendicitis as the primary diagnosis), perforation rate, in-hospital death rate and length of stay.

Results: The diagnostic accuracy rate among the male patients was stable throughout the decade; among the female patients it rose significantly, from 71.7% in 1981 to 75.3% in 1991 ($p < 0.01$). The perforation rates increased significantly among both the female and male patients ($p < 0.01$), whereas the mean length of stay decreased ($p < 0.05$). Despite sex-related differences in the accuracy rates, the male and female patients had similar in-hospital death rates and mean lengths of stay. The institutional diagnostic accuracy rates, as determined from data for 1989–90 to 1991–92, ranged from 50.0% to 96.7%. Multivariate analyses of 27 189 confirmed cases of appendicitis at 175 hospitals revealed that perforation was a strong predictor of in-hospital death (odds ratio [OR] 2.46, 95% confidence interval [CI] 1.24 to 4.88), but comorbidity was the strongest predictor (OR 11.50, 95% CI 5.96 to 22.10). For each 10% increase in the diagnostic accuracy rate, the perforation rate increased 14% (OR 1.14, 95% CI 1.10 to 1.19), but the accuracy rate was not significantly related to the in-hospital death rate or the length of stay.

Conclusion: A higher diagnostic accuracy rate is associated with more perforated appendixes. Although perforation itself leads to adverse outcomes, a higher accuracy rate does not. This suggests that hospitals with higher accuracy rates incur more perforations, but, with close observation, timely laparotomy and the use of modern antibiotics, these patients have favourable outcomes. This contrasts with adverse effects of perforation among patients at high risk for perforation (especially very young children and elderly people) in centres at all accuracy levels. The variation in hospitals' diagnostic accuracy rates suggests that some proportion of appendectomies could be safely avoided.

Objectif : Vérifier l'hypothèse selon laquelle, avec des méthodes modernes de diagnostic et des antibiotiques, le recours plus conservateur aux interventions chirurgicales dans des cas d'appendicite soupçonnée n'entraînerait pas une hausse des taux de complications à court terme dans des cas confirmés.

Drs. Wen and Naylor are with the Institute for Clinical Evaluative Sciences in Ontario and the Department of Preventive Medicine and Biostatistics, University of Toronto, Toronto, Ont. Dr. Naylor is also with the departments of Medicine and Surgery, the Clinical Epidemiology and Health Care Research Program (Sunnybrook Unit) and the Graduate Department of Community Health, University of Toronto, Toronto, Ont.

***Reprint requests to:** Dr. C. David Naylor, Institute for Clinical Evaluative Sciences in Ontario, Rm. G106, 2075 Bayview Ave., North York ON M4N 3M5*

← For prescribing information see page 1705

Conception : Étude stratifiée par observation fondée sur des données administratives.

Contexte : Tous les hôpitaux de l'Ontario où l'on a procédé à des appendicectomies primaires entre le 1er avril 1981 et le 31 mars 1992.

Patients : Les 126 815 patients hospitalisés pour une appendicectomie primaire au cours de la période d'étude.

Mesures des résultats : Taux d'exactitude du diagnostic (appendicite aiguë comme diagnostic primaire), taux de perforation, taux de mortalité à l'hôpital et durée de l'hospitalisation.

Résultats : Le taux d'exactitude du diagnostic chez les patients de sexe masculin est demeuré stable pendant toute la décennie; chez les patientes, il a augmenté considérablement pour passer de 71,7 % en 1981 à 75,3 % en 1991 ($p < 0{,}01$). Les taux de perforation ont augmenté considérablement chez les patientes et les patients ($p < 0{,}01$), tandis que la durée moyenne du séjour a diminué ($p < 0{,}05$). Malgré les écarts liés au sexe au niveau des taux d'exactitude, les taux de mortalité à l'hôpital et la durée moyenne du séjour étaient semblables chez les patients et les patientes. Les taux d'exactitude des diagnostics posés en établissement, établis à partir des données de 1989–1990 à 1991–1992, ont varié de 50,0 % à 96,7 %. Des analyses multivariées de 27 189 cas confirmés d'appendicite dans 175 hôpitaux ont révélé que la perforation était un solide prédicteur de mort à l'hôpital (ratio des probabilités [RP] de 2,46, intervalle de confiance à 95 % [IC] de 1,24 à 4,88), mais la comorbidité était le prédicteur le plus solide (RP de 11,50, IC à 95 % de 5,96 à 22,10). Pour chaque hausse de 10 % du taux d'exactitude du diagnostic, le taux de perforation a augmenté de 14 % (RP de 1,14, IC à 95 % de 1,10 à 1,19), mais il n'y avait aucun lien important entre le taux d'exactitude et le taux de mortalité à l'hôpital ou la durée du séjour.

Conclusion : On établit un lien entre un taux plus élevé d'exactitude du diagnostic et l'augmentation des appendices perforés. Même si la perforation entraîne en soi des résultats défavorables, un taux d'exactitude plus élevé n'en entraîne pas. Cela indique que dans les hôpitaux où les taux d'exactitude sont plus élevés, les perforations sont plus nombreuses. Cependant, après une observation rapprochée, une laparotomie rapide et l'utilisation d'antibiotiques modernes, ces patients présentent de meilleurs résultats. On peut comparer ces résultats aux effets défavorables de la perforation chez les patients à risque élevé de perforation (surtout les enfants très jeunes et les personnes âgées) dans les centres à tous les niveaux d'exactitude. La variation des taux d'exactitude du diagnostic entre les hôpitaux indique que l'on pourrait éviter sans danger un certain pourcentage d'appendicectomies.

Accurate diagnosis and appendectomy remain the cornerstones of therapy for acute appendicitis, but clinical diagnosis can be difficult. A trade-off is widely assumed to exist.[1–5] Delay in laparotomy to improve diagnostic certainty risks organ perforation and sepsis. Velanovich and Satava[2] reviewed 18 clinical studies published up to 1991 and found a positive relation between diagnostic accuracy and perforation. They did not examine the relation between accuracy and mortality or morbidity directly; instead, they inferred a relation indirectly from that between diagnostic accuracy and perforation and that between perforation and mortality or morbidity. They first calculated the average mortality and complication rates among patients with a normal appendix, those with acute appendicitis and those with a perforated appendix; they then applied these rates to their decision-analysis models, with the assumption that if improved diagnostic accuracy resulted in more perforations the mortality and morbidity rates would necessarily increase along with the perforation rate.[2] Based on this indirect inference, they suggested a more aggressive surgical approach in cases of suspected appendicitis. Aggressive surgery, however, risks the removal of many normal appendixes (negative primary appendectomy), with attendant morbidity and expense.[6]

Thus, agreement and evidence are lacking on the appropriate negative appendectomy rate for current practice. We accordingly undertook a study to determine to what extent a higher diagnostic accuracy rate, and therefore a reduction in the number of normal appendixes being removed, would lead to short-term adverse outcomes. We hypothesized that with the availability of abdominal ultrasonography and laparoscopy for diagnosis and broad-spectrum antibiotics to mitigate harm from perforation, a more conservative use of surgery in cases of suspected appendicitis would lead to higher diagnostic accuracy rates without increasing short-term adverse outcomes among confirmed cases.

We used discharge abstracts from 1981 to 1991 for all general hospitals in Ontario. Unlike earlier studies using hospital discharge data,[7,8] our focus was on comparing short-term outcomes across population subgroups and hospitals with different diagnostic accuracy levels. From a descriptive, longitudinal perspective we first compared temporal trends in accuracy rates and corresponding perforation rates, lengths of stay and in-hospital death rates by sex. We then performed formal multivariate analyses to determine the relation between hospital accuracy rates and patient outcomes in the last 3 years under study.

Part of the material presented in this article was published in a limited-circulation monograph.[9] The clinical implications of the findings, however, suggested the need for review by, and publication in, a general medical journal. In this study we provide further analytic detail and discussion and explore the use of alternative multivariate models.

METHODS

DATA SOURCES AND DEFINITIONS

Discharge abstracts for all acute care hospital separations (discharges, transfers or in-hospital deaths) in Ontario are entered into a computer database by a nonprofit institute (the Canadian Institute for Health Information [formerly the Hospital Medical Records Institute]). Information from discharge summaries, operative notes and pathology reports is coded by qualified technicians, who follow the clinical modification of the ninth revision of the International Classification of Diseases (ICD-9-CM)[10] to code diagnoses and the Canadian Classification of Diagnostic, Therapeutic and Surgical Procedures[11] to code procedures.

Records were selected for all Ontario residents with full demographic data who underwent a primary appendectomy from Apr. 1, 1981, to Mar. 31, 1992. For cases to be considered a positive primary appendectomy, the ICD-9-CM code for acute appendicitis or for appendicitis, unqualified, was required; concurrent perforation codes were noted. Negative primary appendectomy was noted if the appendectomy was the primary procedure and the primary diagnosis was unrelated to the appendix. The rate of diagnostic accuracy was simply the proportion of appendectomies that were positive. Because of ambiguity about diagnostic criteria and clinical implications, patients who underwent appendectomy for nonacute appendiceal conditions (including subacute, chronic and relapsing appendicitis) were excluded. Relevant procedure and diagnostic codes are shown in Appendix 1.

Information on the patient's age and sex, the date of admission to hospital, in-hospital death and length of stay in hospital were incorporated as needed. A comorbidity index for each patient was calculated with the use of an ICD-9-CM-based adaptation[12] of a system described by Charlson and associates.[13] In studies of treatment outcomes, one potential confounding factor is the difference in rates of secondary or pre-existing diseases. The comorbidity index is designed to reflect these other conditions so that adjustments can be made in comparing outcomes. In our study sample, comorbid conditions were identified from up to seven additional discharge diagnoses (2nd to 8th diagnosis); the first diagnosis was used to define appendicitis status. Ontario discharge abstracts permit standard calculation of this comorbidity index in nearly all cases.[14]

To reduce the outlier effect, we used the 98th percentile trim points for length of stay in complicated and uncomplicated cases of appendectomy, as defined by the Canadian Institute for Health Information. Thus, length of stay was truncated at 9 days in uncomplicated cases and at 19 days in complicated cases. Information on the number of beds and the teaching status (presence or absence of house staff) for each hospital was taken from the *Canadian Hospital Directory, 1991–1992*.[15]

STATISTICAL ANALYSIS

Sex-specific temporal trends were determined for diagnostic accuracy rates, perforation rates, in-hospital death rates and lengths of stay. Analyses at the hospital level were confined to the 3 most recent fiscal years (1989–90 to 1991–92) because of temporal shifts in accuracy rates and technology. To frame the hospital analyses, first we compared patient outcomes after grouping hospitals according to institutional diagnostic accuracy rates (hospitals with rates below the 25th percentile, those with rates from the 25th to the 75th percentile, and those with rates above the 75th percentile). Then, we used logistic and linear regression analyses to examine the independent effect of the diagnostic accuracy rate on patient outcomes (perforation rate, in-hospital death rate and length of stay) after adjusting for other factors, including the admitting hospital's appendectomy volume, number of beds and teaching status, and the patient's sex, age and comorbidity index. Each patient was assigned the values pertaining to the hospital to which he or she was admitted. Regression analyses were repeated after excluding patients admitted to low-volume hospitals (those with fewer than 20 primary appendectomies during the 3 fiscal years under study).

Because the combination of variables pertaining to the hospital and to the patient may affect the variance of the independent variables in the regression models and cause multicollinearity, we also performed multiple linear regression analyses involving independent variables at the hospital level. In these, hospital-specific means or proportions were calculated for patient characteristics (e.g., age, sex, comorbidity, perforation, in-hospital death and length of stay) and entered as independent variables into the linear regression models. Since hospital-specific means or proportions were based on different sample sizes, we applied a weighted least squares regression method, with primary appendectomy volume as a weight, to reflect the relative contribution of each hospital to the model.[16]

RESULTS

OVERALL PROFILE

During the decade under study 103 768 positive primary appendectomies and 23 047 negative primary appendectomies were performed in Ontario hospitals; 6249 patients with diagnoses related to the appendix other than acute appendicitis were excluded. The overall diagnostic accuracy rate was 81.8% (73.9% among the female patients and 89.0% among the male patients). The perforation rate was 20.6%, the in-hospital death rate 0.1% and the mean length of stay 5.3 days.

The most frequently recorded primary diagnoses for negative primary appendectomies were abdominal pain and mesenteric lymphadenitis among the male patients and abdominal pain, mesenteric lymphadenitis and gynecologic conditions among the female patients (a complete list of all the conditions is available from the corresponding author upon request).

Temporal trends and subgroup comparisons

The diagnostic accuracy rate increased during the study decade among the female patients, from 71.7% in 1981–82 to 75.3% in 1991–92; there was no increase among the male patients (Table 1). The perforation rate increased among both the male and female patients, but the length of stay decreased. No temporal changes in the in-hospital death rates were observed for either sex (Table 1).

In the 3 fiscal years studied for the hospital-level analyses, there were 27 189 positive and 5975 negative primary appendectomies; 1343 appendectomies performed because of chronic appendiceal diseases were excluded. As expected given diagnostic confusion from gynecologic disorders, the diagnostic accuracy rates were significantly lower among the female patients aged 15 to 64 years than among the male patients in the same age groups (Table 2). The perforation rates also were lower among the female patients in these age groups, but the differences were generally much smaller than those for the accuracy rates. The mean lengths of stay were similar for the male and female patients. Only 1 of the 23 142 patients less than 45 years old with confirmed appendicitis died postoperatively.

Hospital-level cross-tabulation

After 13 low-volume hospitals with extreme values were excluded, the institutional diagnostic accuracy rate was found to range from 50.0% to 96.7%. The in-hospital death rates and the mean lengths of stay were similar for hospitals with different levels of diagnostic accuracy (Table 3), but the perforation rates were slightly higher for the hospitals with higher diagnostic accuracy rates. The proportions of female patients and of older patients were lower in the hospitals with higher diagnostic accuracy rates; this observation highlighted the need for a multivariate analysis.

Multivariate analysis

In the multivariate analysis of individual patient data

Table 1: Diagnostic accuracy rates, perforation rates, in-hospital death rates and lengths of stay (LOS) in hospital among patients in Ontario who underwent appendectomy for suspected acute appendicitis from Apr. 1, 1981, to Mar. 31, 1992, by sex

	Female				Male			
Year	Accuracy rate, %	Perforation rate, %	Death rate, %	Mean LOS (and SD*), d	Accuracy rate, %	Perforation rate, %	Death rate, %	Mean LOS (and SD), d
1981	71.7	17.5	0.09	5.7 (3.3)	89.6	20.7	0.15	5.6 (3.4)
1982	72.7	17.0	0.11	5.6 (3.2)	89.0	21.0	0.17	5.4 (3.3)
1983	74.0	17.8	0.12	5.4 (3.2)	88.9	21.3	0.07	5.4 (3.3)
1984	73.3	18.2	0.22	5.4 (3.2)	88.6	19.7	0.06	5.3 (3.2)
1985	74.0	15.8	0.00	5.3 (3.4)	89.0	19.6	0.06	5.3 (3.6)
1986	74.9	19.3	0.05	5.5 (3.6)	88.7	21.6	0.17	5.3 (3.6)
1987	75.3	19.1	0.02	5.4 (3.5)	88.9	21.5	0.04	5.3 (3.6)
1988	75.2	18.7	0.09	5.3 (3.5)	90.0	22.8	0.11	5.2 (3.5)
1989	73.0	22.9	0.14	5.2 (4.4)	88.7	22.1	0.10	5.4 (4.3)
1990	73.5	22.2	0.08	5.5 (4.7)	89.3	25.3	0.16	5.2 (4.0)
1991	75.3	21.6	0.21	4.7 (3.1)	88.8	23.8	0.17	4.5 (3.0)
Slope†	0.21‡	0.44‡	0.00§	−0.06‖	−0.01§	0.35‡	0.01§	−0.06‖

*SD = standard deviation.
†Simple linear regression coefficient for continuous variables, and gradient in proportions for dichotomous variables, by year.[17]
‡$p < 0.01$.
§Not significant.
‖$p < 0.05$.

Table 2: Diagnostic accuracy rates and treatment outcomes among 33 164 patients in Ontario who underwent appendectomy for suspected acute appendicitis from Apr. 1, 1989, to Mar. 31, 1992, by age group and sex*

Age group, yr; sex†	Diagnostic accuracy rate, %	No. (and %) of patients with perforation	No. (and %) of patients who died in hospital	Mean LOS (and SD), d
≤ 14				
Female (2 376)	83.1	598 (25.2)	1 (0.0)	4.8 (3.4)
Male (3 453)	89.9‡	835 (24.2)	0	4.6 (3.4)
15–29				
Female (4 775)	72.3	612 (12.8)	0	4.6 (3.4)
Male (6 524)	91.3‡	1 125 (17.2)‡	0	4.3 (3.1)
30–44				
Female (2 437)	70.6	534 (21.9)	0	5.3 (4.0)
Male (3 577)	89.4‡	829 (23.2)	0	5.2 (3.9)
45–64				
Female (1 215)	74.4	407 (33.5)	2 (0.2)	6.9 (4.7)
Male (1 706)	85.4‡	633 (37.1)§	4 (0.2)	6.8 (4.7)
65–74				
Female (341)	71.3	147 (43.1)	4 (1.2)	9.3 (5.5)
Male (396)	72.1	192 (48.5)	4 (1.0)	9.2 (5.4)
≥ 75				
Female (185)	61.7	107 (57.8)	9 (4.9)	11.8 (5.5)
Male (204)	66.9	105 (51.5)	15 (7.4)	11.0 (5.9)

*Outcome assessment was confined to the 27 189 patients with confirmed appendicitis at surgery (positive appendectomy).
†Numbers in parentheses represent patients who had a positive appendectomy.
‡$p < 0.01$.
§$p < 0.05$.

Table 3: Factors associated with low, medium and high institutional diagnostic accuracy rates among 162 hospitals in Ontario from Apr. 1, 1989, to Mar. 31, 1992*

Factor	Accuracy rate† Low $n = 40$	Medium $n = 81$	High $n = 41$	p value‡
Mean primary appendectomy volume (and SD)	224 (187)	240 (184)	148 (152)	< 0.05
Mean no. of beds (and SD)	159 (154)	154 (126)	93 (85)	< 0.05
Teaching hospital, %	30.0	34.6	17.1	NS
% of patients who were female	43.1	41.6	40.2	< 0.01
Age group, yr; % of patients				< 0.01
≤ 14	19.0	20.8	26.1	
15–29	42.7	42.0	38.8	
30–44	23.2	22.1	20.8	
≥ 45	15.1	15.0	14.3	
% of patients with comorbidity	2.4	2.4	2.6	NS
Perforation rate, %	21.2	22.7	23.5	< 0.01
In-hospital death rate, %	0.2	0.1	0.1	NS
Mean LOS (and SD), d	5.3 (4.0)	5.1 (3.9)	5.1 (3.9)	NS

*Thirteen hospitals with low volumes of either all positive or all negative appendectomies are excluded. Information on primary appendectomy volume, number of beds (acute medical and surgical beds) and teaching status were obtained from the database of the Hospital Medical Records Institute (now the Canadian Institute for Health Information) and published sources (see Methods section). Information on sex, age, comorbidity, perforation, death and LOS were determined for patients admitted to each hospital.
†Low = ≤ 77.50%, medium = 77.51%–87.50%, high = > 87.50%.
‡NS = not significant.

with hospital covariates, perforation was first incorporated as one of the independent variables (Table 4). Perforation was a strong predictor of in-hospital death and a prolonged hospital stay but was markedly weaker than comorbidity as a predictor of in-hospital death.

In another multivariate model the perforation rate was considered as an outcome measure. The relative odds of perforation rose 14% for each 10% absolute increase in the hospital diagnostic accuracy rate (Table 5). That is, as the overall diagnostic accuracy rate rose from, say, the current mean of 82% to a new level of 92%, the perforation rate would be expected to rise from 22.5% to 24.9%. (This derives from the standard conversion for odds ratios, in which the new rate = $0.225 \times 1.14/[1 - 0.225] + [0.225 \times 1.14]$.) Thus, for every 100 operations, avoiding two or three cases of perforation demanded that 10 additional negative appendectomies be performed. However, in neither model did the diagnostic accuracy rate significantly alter the odds ratio for in-hospital death or the coefficient for length of stay.

The number of hospital beds was unrelated to treatment outcomes (Tables 4 and 5), but the appendectomy volume correlated with a shorter length of stay. Similar results were obtained from "ecologic" regression analyses at the hospital level (Tables 6 and 7), although they tended to be nonsignificant, especially for independent variables originally measured at the patient level. The exclusion of patients admitted to low-volume hospitals did not change the results (data not shown).

Discussion

Advances in imaging, the availability of laparoscopy and the effectiveness of newer broad-spectrum antibiotics led us to question the time-honoured clinical axiom that a reduction in the proportion of negative appendectomies (i.e., an increase in the preoperative diagnostic accuracy rate) causes an increase in the perforation rate among patients who actually have appendicitis, which in

Table 4: Odds ratios (ORs) for in-hospital death rates and linear regression coefficients for LOS, from Apr. 1, 1989, to Mar. 31, 1992

	Outcome	
Independent variable	In-hospital death rate, OR (and 95% CI*)	LOS, d; coefficient (and 95% CI)
Diagnostic accuracy rate (10% increase)	0.96 (0.45 to 1.52)	−0.04 (−0.10 to 0.02)
Primary appendectomy volume (50-case increase)	0.98 (0.89 to 1.08)	−0.06 (−0.05 to −0.07)
No. of beds (30-bed increase)	1.00 (0.93 to 1.09)	0.01 (0.00 to 0.02)
Treatment in teaching hospital (no = 0, yes = 1)	1.85 (0.88 to 4.00)	0.36 (0.26 to 0.46)
Age (10-yr increase)	–†	0.56 (0.53 to 0.59)
Sex (female = 0, male = 1)	0.98 (0.50 to 1.91)	−0.30 (−0.21 to −0.39)
Comorbidity (none = 0, any = 1)	11.50 (5.96 to 22.10)	1.96 (1.67 to 2.25)
Perforation (no = 0, yes = 1)	2.46 (1.24 to 4.88)	1.96 (1.85 to 2.07)

*CI = confidence interval.
†OR was calculated only for subjects aged 45 years and over.

Table 5: Odds ratios for perforation and in-hospital death rates and linear regression coefficients for LOS, from Apr. 1, 1989, to Mar. 31, 1992

	Outcome		
Independent variable	Perforation rate, OR (and 95% CI)	In-hospital death rate, OR (and 95% CI)	LOS, d; coefficient (and 95% CI)
Diagnostic accuracy rate (10% increase)	1.14 (1.10 to 1.19)	0.91 (0.58 to 1.43)	0.01 (−0.05 to 0.07)
Primary appendectomy volume (50-case increase)	1.01 (0.98 to 1.02)	0.98 (0.90 to 1.08)	−0.05 (−0.04 to −0.06)
No. of beds (30-bed increase)	0.98 (0.97 to 0.99)	1.01 (0.94 to 1.09)	0.00 (0.00 to 0.01)
Treatment in teaching hospital (no = 0, yes = 1)	1.37 (1.30 to 1.47)	1.79 (0.85 to 3.70)	0.47 (0.37 to 0.57)
Age (10-yr increase)	1.21 (1.19 to 1.23)	–*	0.63 (0.60 to 0.66)
Sex (female = 0, male = 1)	1.14 (1.07 to 1.21)	0.87 (0.46 to 1.66)	−0.25 (−0.16 to −0.34)
Comorbidity (none = 0, any = 1)	1.47 (1.25 to 1.74)	11.30 (6.01 to 21.33)	2.13 (1.83 to 2.43)

*OR was calculated only for subjects aged 45 years and over.

turn leads to a higher in-hospital death rate and prolonged lengths of stay owing to abdominal contamination and sepsis.[1,2] Our findings conflicted with conventional wisdom.

The diagnostic accuracy rates among the female patients in our study rose over time, as did the perforation rates. Among the male patients the diagnostic accuracy rates remained stable, but the perforation rates rose. Comparisons across subgroups by age and sex among 33 164 patients in the fiscal years 1989–90 to 1991–92 also failed to confirm the expected relations. The diagnostic accuracy rates were significantly lower among the female patients aged 15 to 64 years than among the male patients in these age groups, presumably because of diagnostic uncertainty caused by gynecologic conditions. The perforation rates tended to be lower among the female patients in these age groups as well. However, there were no sex-related differences in the length of stay, and in-hospital deaths were extremely rare among the younger patients. There was definite variation

Table 6: Linear regression coefficients for in-hospital death rates and mean LOS, from Apr. 1, 1989, to Mar. 31, 1992, at the hospital level

	Outcome; coefficient (and 95% CI)	
Independent variable	In-hospital death rate	LOS, d
Diagnostic accuracy rate (10% increase)	−0.0182 (−0.0786 to 0.0422)	−0.09 (−0.26 to 0.09)
Primary appendectomy volume (50-case increase)	−0.0008 (−0.0137 to 0.0121)	−0.07 (−0.03 to −0.10)
No. of beds (30-bed increase)	−0.0009 (−0.0144 to 0.0126)	0.03 (−0.00 to 0.07)
Teaching hospital (no = 0, yes = 1)	0.0524 (−0.0456 to 0.1504)	0.36 (0.08 to 0.64)
Hospital-specific mean patient age (10-yr increase)	0.1056 (0.0066 to 0.2046)	0.06 (−0.22 to 0.35)
Hospital-specific sex ratio* (1% increase)	0.0062 (−0.0011 to 0.0135)	0.01 (−0.01 to 0.03)
Hospital-specific comorbidity rate† (1% increase)	0.0275 (−0.0003 to 0.0553)	0.05 (−0.03 to 0.13)
Hospital-specific perforation rate‡ (1% increase)	0.0032 (−0.0012 to 0.0086)	0.01 (0.00 to 0.03)

*Percentage of males.
†Absolute percentage of patients with comorbid conditions.
‡Absolute percentage of patients with perforation.

Table 7: Linear regression coefficients for perforation rates, in-hospital death rates and mean LOS, from Apr. 1, 1989, to Mar. 31, 1992, at the hospital level

	Outcome; coefficient (and 95% CI)		
Independent variable	Perforation rate	In-hospital death rate	Mean LOS, d
Diagnostic accuracy rate (10% increase)	1.82 (−0.12 to 3.76)	−0.0114 (−0.0714 to 0.0486)	−0.06 (−0.24 to 0.11)
Primary appendectomy volume (50-case increase)	0.02 (−0.40 to 0.44)	−0.0007 (−0.0136 to 0.0122)	−0.07 (−0.03 to −0.11)
No. of beds (30-bed increase)	−0.16 (−0.60 to 0.28)	−0.0016 (−0.0151 to 0.0119)	0.03 (−0.00 to 0.07)
Teaching hospital (no = 1, yes = 1)	5.32 (2.25 to 8.39)	0.0723 (−0.0228 to 0.1674)	0.44 (0.16 to 0.71)
Hospital-specific mean patient age (10-yr increase)	−0.89 (−4.01 to 2.32)	0.1023 (0.0029 to 0.2017)	0.05 (−0.24 to 0.34)
Hospital-specific sex ratio* (1% increase)	0.08 (−0.16 to 0.32)	0.0065 (−0.0008 to 0.0138)	0.01 (−0.01 to 0.03)
Hospital-specific comorbidity rate* (1% increase)	0.45 (−0.44 to 1.35)	0.0291 (0.0013 to 0.0569)	0.06 (−0.03 to 0.14)

*As defined in Table 6.

among the hospitals in the rates of removal of normal appendixes. When hospitals were grouped by institutional accuracy levels, those in the upper 25th percentile had minimally higher perforation rates than the other hospitals but similar lengths of stay and in-hospital death rates. These "ecologic" data provided prima facie evidence that, despite an association with minor increases in perforation rates, higher diagnostic accuracy rates do not lead to increases in morbidity or mortality.

We then analysed the relation between the institutional diagnostic accuracy rates and outcomes at the patient level using multivariate techniques and including data on 27 189 positive primary appendectomies in 175 hospitals. When considered as a determinant of outcomes, perforation was associated with an increased risk of in-hospital death and prolonged length of stay. However, the limited overall impact of perforation was highlighted by perforation rates of 30% to 50% among patients over 45 years of age, with in-hospital death rates above 1% only among the few patients over 75 years old who underwent surgery for suspected appendicitis. By far the strongest determinant of death was comorbidity. Whether perforation was introduced as a determinant or considered as an outcome itself, the diagnostic accuracy rate was found to be unrelated to an increased risk of in-hospital death or prolonged length of stay.

Our regression models appeared sound, with estimated effects for important prognostic variables (e.g., age and comorbidity) that were in the right direction and of the expected magnitude. As with any study of the relation between diagnostic accuracy and clinical outcomes, accuracy had to be considered in aggregate at the hospital level. Use of such "ecologic" covariates in an analysis of individuals will lead to underestimation of the variance of independent variables in the model.[18] However, a parallel ecologic regression analysis, which aggregated variables at the individual level for each hospital, showed similar relations between the diagnostic accuracy rate and various outcomes. It therefore seems unlikely that the mixing of ecologic and individual variables confounded the analysis.

Moreover, the enforcement of a "homogeneous" level in regression variables through the use of hospital-specific means or proportions instead of individual data has its own shortcomings. The precision of any estimates of effects for variables at the individual level is greatly reduced; we are left with a model in which the "n" is the number of hospitals and nothing more. Many clinically meaningful and strong outcome predictors, such as age, sex and comorbidity, became nonsignificant in our ecologic models because of the extreme compression of the range for these variables and the loss of precision in their estimates of effect size when hospital-specific means or proportions were calculated. Indeed, underestimation of the variances in the model by our combining ecologic and individual variables is, if anything, desirable here, since we sought to rule out any potential adverse outcomes from rising diagnostic accuracy rates. Statistical power to detect a relation to in-hospital death rates remained very low, since only 1 of 23 142 patients under 45 years of age died and only 38 of 4047 patients 45 and over died. However, these findings sharply limit the magnitude of any absolute risk gradient that could occur with rising perforation rates and affirm the safety of a conservative approach to surgery for suspected appendicitis, especially among people under 45.

Our sample sizes compared favourably with the total of 10 023 appendectomies in the 18 clinical studies reviewed by Velanovich and Satava.[2] They used their literature review to generate a decision analysis, highlighting the trade-offs between lowering the perforation rate and raising the negative appendectomy rate. They concluded that "an aggressive approach to patients with suspected appendicitis is warranted . . . because most of the mortality and morbidity from appendectomy is related to perforated appendicitis." However, as noted in our introduction, their analysis lacked a direct assessment of the relation between diagnostic accuracy rates and mortality or morbidity and a multivariate framework, including explicit consideration of comorbidity and age–sex subgroup data, as presented here; it also incorporated data from older case series.

We believe that our findings can be readily reconciled with clinical thinking if two categories of perforation are considered. The first is exemplified by the person who arrives for medical care with established perforation and generalized sepsis. Elderly people (and very young people) appear especially prone to perforation early in the course of appendicitis;[3] they are also patients for whom the diagnosis is obscure and sometimes overlooked. It is not surprising that perforation will be associated with delayed discharge and occasional death among these patients. The second category is exemplified by the person with abdominal pain due to evolving appendicitis. This archetypal individual will be between 15 and 64 years of age and will have no perforation at the time of seeking medical care. Watchful waiting or diagnostic imaging in hospital, with rapid surgery once the diagnosis is clarified, may lead to higher diagnostic accuracy and perforation rates. However, with modern antibiotics and no prolonged period of frank perforation and abdominal contamination, the outlook for most of these patients matches that of patients with appendicitis but no perforation at laparotomy. This view is supported by primary data showing that the negative appendectomy rate falls with increasing time between presentation and laparotomy, without an increase in complication rates.[19]

The institutional rates of diagnostic accuracy varied substantially in our analysis. Increased appendectomy volume at the hospital level had a statistically significant impact in reducing length of stay and a positive, although nonsignificant, impact on the in-hospital death rate; these findings suggest that the "practice-makes-perfect" factor seen with major surgery[20] is partly applicable even for a common procedure like appendectomy. The perforation rates were higher and the lengths of stay longer in the teaching hospitals (Tables 4 and 5); a weak trend to higher in-hospital death rates in the teaching hospitals was also noted. These observations may reflect the influences of both inexperienced house staff and unmeasured referral biases not eliminated by adjustments for age and comorbidity.

The issue of unmeasured referral biases underscores the fact that any study drawing on administrative data necessarily suffers from a lack of clinical detail and is subject to vagaries in the coding of diagnoses and procedures. A particular concern is that the clinical diagnosis of presumed acute appendicitis may be accepted in the absence of pathological confirmation if the medical records staff do not seek a pathological report in the hospital chart, or if the report is delayed. The observed increase in the perforation rates over the study period, for example, could reflect greater assiduousness on the part of the medical records staff, a self-critical attitude by surgeons that leads to a lower threshold for noting possible perforation in the operative note or even an indirect influence of a differential fee for cases of appendicitis with perforation. Fortunately, concerns about temporal coding shifts do not apply to our multivariate analyses, which drew on aggregate data for the fiscal years 1989–90 to 1991–92. We are also unable to generate a plausible hypothesis that allows coherent attribution of our multivariate findings to systematic coding errors. Perhaps the only alternative unifying hypothesis would be that the diagnostic accuracy rates and the perforation rates primarily reflect the assiduousness of the pathology department, which may be unrelated to surgical skill. If so, the grossest of perforations are those uniformly captured at surgery, with the result that perforation is associated with death, but relations between perforation, diagnostic accuracy and in-hospital death rates are otherwise obscured by variations in pathology reporting, practices and coding errors.

Another potential limitation is the lack of information on long-term outcomes. There is some controversy about whether appendicitis with perforation in women leads to an increased risk of infertility from tubal dysfunction. Mueller and collaborators,[21] for example, suggested that women with a perforated appendix are three to five times as likely to have tubal infertility as women who have never had appendicitis. However, the perforations in question may have been remote, and the same risks may not apply with current surgical management and antibiotic therapy. Also, their risk estimates were drawn from a case–control study in which only 5% to 6% of the patients with primary or secondary tubal infertility recalled having had a "ruptured appendix," as compared with 1% to 2% of the control subjects. Clearly, most cases of tubal infertility are unrelated to surgical management of appendiceal disease. An aggressive surgical approach based on fears of perforation and tubal damage would cause a great many women to undergo negative appendectomies for every one who might conceivably be spared tubal infertility.

As Koepsell, Inui and Farewell[3] and Chang, Hogle and Welling[6] argued, perforation alone is not a good indicator of the quality of surgical care, especially when one considers the burden on the patient and the cost to the hospital associated with negative appendectomy. We were unable to establish a particular threshold at which the rising diagnostic accuracy rate was associated with an increasing length of stay and higher in-hospital death rate. We do not know why some of the hospitals had higher accuracy rates than others; the only hospital-level predictors of a higher accuracy rate and shorter length of stay were identified earlier. However, up to 900 negative primary appendectomies could be safely avoided in Ontario each year if the diagnostic accuracy rate in all hospitals below the 75th percentile could be increased to the 75th percentile (i.e., over 87.5%).

In summary, our findings are consistent with the concept that very good standards of care have already been attained and that further improvements in preoperative diagnosis of appendicitis are possible in many hospitals, without harm to patients. Recent dissemination of hospital-specific data on diagnostic accuracy rates for Ontario may prompt and facilitate local audits.[9] For hospitals with a low diagnostic accuracy rate confirmed by chart review, improvements may be possible through the use of diagnostic algorithms[22,23] or increased reliance on watchful waiting, as was proven effective 20 years ago.[24] Ultrasonography and laparoscopy are also useful tools for the differential diagnosis. For example, two recent randomized trials have demonstrated that the use of diagnostic laparoscopy among patients with suspected appendicitis substantially reduces the rate of unnecessary appendectomy without compromising treatment outcomes.[25,26] Supported by modern diagnostic methods, the above-mentioned strategies may be safe and effective alternatives to early exploratory surgery, especially among people aged 15 to 45 years, who constitute most patients undergoing appendectomy.

We thank Marc-Erick Theriault for computer programming, Caitlin Davies for administrative support, Drs. Brian Kenney and Marsha Cohen for related collaborations in this area and Dr. Jack Williams for critical appraisal of the manuscript. We also thank the following surgical colleagues for their reviews and comments: Drs. Phillip T. Barron, Robin J. Fairfull-Smith, Duncan M. Paterson and Brian M. Taylor.

This study was supported by the Institute for Clinical Evaluative Sciences in Ontario. Dr. Naylor is the recipient of a Career Scientist Award from the Ontario Ministry of Health.

References

1. Pass HI, Hardy JD: The appendix. In Hardy JD (ed): *Hardy's Textbook of Surgery*, 2nd ed, JB Lippincott, Philadelphia, 1988: 574–581

2. Velanovich V, Satava R: Balancing the normal appendectomy rate with the perforated appendicitis rate: implications for quality assurance. *Am Surg* 1992; 58: 264–269

3. Koepsell TD, Inui TS, Farewell VT: Factors affecting perforation in acute appendicitis. *Surg Gynecol Obstet* 1981; 153: 508–510

4. Dunn EL, Moore EE, Elerding SC et al: The unnecessary laparotomy for appendicitis — Can it be decreased? *Am Surg* 1982; 48: 320–323

5. Anderson RE, Hugander A, Thulin AJG: Diagnostic accuracy and perforation rate in appendicitis: association with age and sex of the patient and with appendectomy rate. *Eur J Surg* 1992; 158: 37–41

6. Chang FC, Hogle HH, Welling DR: The fate of the negative appendix. *Am J Surg* 1973; 126: 752–754

7. Addiss DG, Shafter N, Fowler BS et al: The epidemiology of appendicitis and appendectomy in the United States. *Am J Epidemiol* 1990; 132: 910–925

8. Luckmann R: Incidence and case fatality rates for acute appendicitis in California. *Am J Epidemiol* 1989; 129: 905–918

9. Wen SW, Naylor CD: Appendectomies in Ontario: outcomes and recent hospital-specific utilization patterns. In Naylor CD, Anderson GM, Goel V (eds): *Patterns of Health Care in Ontario*, Canadian Medical Association, Ottawa, 1994: 143–150

10. US National Center for Health Statistics: *International Classification of Diseases, 9th rev (Clinical Modification)*, vols 1–3, Commission on Professional and Hospital Activities, Ann Arbor, Mich, 1986

11. *Canadian Classification of Diagnostic, Therapeutic, and Surgical Procedures*, Statistics Canada, Ottawa, 1986

12. Deyo RA, Cherkin DC, Ciol MA: Adapting a clinical comorbidity index for use with ICD-9-CM administrative database. *J Clin Epidemiol* 1992; 45: 613–619

13. Charlson ME, Pompei P, Ales KL et al: A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987; 40: 373–383

14. Chen E, Naylor CD: Variation in hospital length of stay for acute myocardial infarction in Ontario, Canada. *Med Care* 1994; 32: 420–435

15. *Canadian Hospital Directory, 1991–1992*, vol 39, Canadian Hospital Association, Ottawa, 1992

16. Afifi AA, Clark V: *Computer-Aided Multivariate Analysis*, Lifetime Learning Publications, Belmont, Calif, 1984

17. Fleiss JL: *Statistical Methods for Rates and Proportions*, 2nd ed, John Wiley & Sons, New York, 1981: 143–146

18. Zeger SL, Liang KY: Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986; 42: 121–130

19. McLean AD, Stonebridge PA, Bradbury AW et al: Time of presentation, time of operation, and unnecessary appendicectomy. *BMJ* 1993; 306: 307

20. Hannan EL, Kilburn H Jr, O'Donnell JF et al: A longitudinal analysis of the relationship between in-hospital mortality in New York State and the volume of abdominal aortic aneurysm surgery performed. *Health Serv Res* 1992; 27: 517–542

21. Mueller BA, Daling JR, Moore DE et al: Appendectomy and the risk of tubal infertility. *N Engl J Med* 1986; 315: 1506–1508

22. Detmer DE, Frish C: Improved results in acute appendicitis care following areawide review. *Med Decis Making* 1984; 4: 217–227

23. Adams ID, Chan M, Clifford PC et al: Computer-aided diagnosis of acute abdominal pain: a multi-centre study. *BMJ* 1986; 293: 800–804

24. White JJ, Santillana M, Haller JA: Intensive in-hospital observation: a safe way to decrease unnecessary appendectomy. *Am Surg* 1975; 41: 793–798

25. Olsen JB, Myren CJ, Hoahr PE: Randomized study of the value of laparoscopy before appendicectomy. *Br J Surg* 1993; 83: 922–923

26. Jadallah FA, Abdul-Ghani AA, Tibbin S: Diagnostic laparoscopy reduces unnecessary appendicectomy in fertile women. *Eur J Surg* 1994; 160: 41–45

Appendix 1: Diagnostic and procedure codes used to define and categorize patients who underwent appendectomy for suspected acute appendicitis in Ontario hospitals between Apr. 1, 1981, and Mar. 31, 1992

Type of appendectomy*	Description	Code type	Codes†
Primary	Appendicitis is primary reason for surgery	Procedure	CCP 59.0
Negative	Primary diagnosis is unrelated to appendix	Diagnosis	Miscellaneous
Positive	Primary diagnosis is acute appendicitis or appendicitis, unqualified	Diagnosis	ICD-9-CM 540.0, 540.1, 540.9 and 541

*Incidental appendectomies (CCP code 59.2) and primary appendectomies with ICD-9-CM codes 542, 543.0 and 543.9 were excluded as nonacute cases.
†CCP = Canadian Classification of Diagnostic, Therapeutic and Surgical Procedures;[11] ICD-9-CM = clinical modification of the 9th revision of the International Classification of Diseases.[10]

For prescribing information see page 1690 →